



Demographics of the Gay and Lesbian Population in the United States: Evidence from Available Systematic Data Sources

Author(s): Dan Black, Gary Gates, Seth Sanders, Lowell Taylor

Source: *Demography*, Vol. 37, No. 2 (May, 2000), pp. 139-154

Published by: Population Association of America

Stable URL: <http://www.jstor.org/stable/2648117>

Accessed: 26/01/2009 20:00

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=paa>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



Population Association of America is collaborating with JSTOR to digitize, preserve and extend access to *Demography*.

<http://www.jstor.org>

DEMOGRAPHICS OF THE GAY AND LESBIAN POPULATION IN THE UNITED STATES: EVIDENCE FROM AVAILABLE SYSTEMATIC DATA SOURCES*

DAN BLACK, GARY GATES, SETH SANDERS, AND LOWELL TAYLOR

This work provides an overview of standard social science data sources that now allow some systematic study of the gay and lesbian population in the United States. For each data source, we consider how sexual orientation can be defined, and we note the potential sample sizes. We give special attention to the important problem of measurement error, especially the extent to which individuals recorded as gay and lesbian are indeed recorded correctly. Our concern is that because gays and lesbians constitute a relatively small fraction of the population, modest measurement problems could lead to serious errors in inference. In examining gays and lesbians in multiple data sets we also achieve a second objective: We provide a set of statistics about this population that is relevant to several current policy debates.

The emergence of solid demographic studies describing the gay and lesbian population marks an important change for social science research. Historically, few sizable surveys of this population were available, and many previous surveys that provided large samples of gays and lesbians utilized “convenience sampling,” as in samples drawn from readers of particular magazines or newspapers, or responses solicited from Internet sites or in gay bars. Researchers have been properly reluctant to draw general inferences about the gay and lesbian population from these samples. Recently, however, a number of scholars have begun to study economic and social issues in the gay and lesbian population using sizable samples with known properties—samples drawn from the General Social Survey, the National Health and Social Life Survey, and the 1990 U.S. census.

We view this recent emergence of careful, systematic empirical work on the gay and lesbian population as valuable on two fronts. First, this work can usefully inform public policy. The past decade has been marked by a significant amount of public debate and legislation regarding gay and lesbian Americans. Issues include initiatives designed to prohibit discrimination or, conversely, to prohibit civil rights protection based on sexual orientation; public policy concern-

ing provision of domestic partnership benefits (including health insurance) to gay and lesbian couples; the U.S. military policy prohibiting openly gay and lesbian individuals from serving in the armed forces; the legalization of same-sex marriage; and gay and lesbian parental rights and suitability for adoption. Informed policy analysis about these issues requires accurate demographic information about the gay and lesbian population. For example, the city of San Francisco was unable to estimate the number of partnered households or the rate at which one domestic partner was not covered by benefits. This situation led Carol Piasente, a spokeswoman for the San Francisco Chamber of Commerce, to comment, “We don’t have a clue about costs. That’s the problem. Nobody knows” (“S.F. Seeks Equal Treatment” 1996:3b).

Second, careful empirical analysis of the gay and lesbian population holds promise for helping social scientists understand a wide array of important questions—questions about the general nature of labor market choices, accumulation of human capital, specialization within households, discrimination, and decisions about geographic location.

Our work provides an overview of available data currently used by social scientists to study the gay and lesbian population in the United States; in doing so, it provides interesting and policy-relevant statistics about this population. We focus on what can be learned about the gay and lesbian population from three large data sets: the General Social Survey (GSS), the National Health and Social Life Survey (NHLS), and the U.S. census. In particular, we examine four characteristics of the gay and lesbian population: their geographical distribution, their veteran status, the family structure of their households, and their education, earnings, and wealth.

We begin by reviewing the limited economic and demographic literature that investigates gays and lesbians using the NHLS, the GSS, or 1990 census data. We then discuss the three data sources, emphasizing how gays and lesbians can be identified in each survey. We pay special attention to confirming that the sample of gays and lesbians identified is not simply the result of recording error. In the next section, we provide a comparison of findings from the data sets (and, in some instances, comparisons with other available information). In doing so, we develop a statistical portrait of the gay and lesbian population that is broadly consistent across data sources. We close with concluding remarks.

LITERATURE REVIEW

The National Health and Social Life Survey (NHLS) served as the basis for two well-known books, *Sex in*

*Dan Black, Center for Policy Research, Syracuse University, Syracuse, NY 13244, dablac01@maxwell.syr.edu. Gary Gates, Heinz School of Public Policy and Management, Carnegie Mellon University; Seth Sanders, Department of Economics, University of Maryland; Lowell Taylor, Heinz School of Public Policy and Management, Carnegie Mellon University. The authors gratefully acknowledge financial support from NICHD Grant HD3703-01 430. We are also grateful to Hoda Makar for research and editorial assistance, and to Jason Fields, Gregory Lewis, Robert Little, Jane Mauldon, Susan Newcomer, and participants in demography workshops at the University of California, Berkeley and the University of Maryland for useful comments.

America: A Definitive Study (Michael et al. 1994) and *The Social Organization of Sex: Sexual Practices in the United States* (Laumann et al. 1994). The latter book features a chapter (Chapter 8) on gays and lesbians that focuses on the definition of homosexuality and the prevalence of gay, lesbian, and bisexual behavior in the United States. One of the main issues addressed by Laumann et al. is how varying definitions of homosexuality affect the measured incidence rates. The authors show that whereas the incidence rate of homosexual desire is 7.7% for men and 7.5% for women, the rate at which men identify themselves as gay is 2.8%, and the rate at which women identify themselves as lesbians is 1.4%. These figures are similar to the rates at which men and women have exclusively same-sex sex (3.0% and 1.6%).

The authors' findings are important for two reasons. First, they demonstrate the importance of sampling from a known population. There exists a widespread belief, based largely on Kinsey's pioneering research (e.g., Kinsey et al. 1948), that "10 percent of males are more or less exclusively homosexual." This statement is not supported in the careful work of Laumann et al. (1994).¹ Second, they highlight the ambiguity of the very definition of homosexuality. Those who acknowledge homosexual desires may be far more numerous than those who actually act on those desires. Nonetheless, gays and lesbians show substantially different behavior than other individuals on some important dimensions: For example, they are less likely to enter into traditional marriages.

In addition to standard economic and demographic data, the NHLS collects by far the most extensive information on sexual practices and sexual partners. Among the data sets we examine, the NHLS is the only data set that gathers information on sexual practices over the life course—an important advantage because sexual behavior regarding the sex of one's partner is not immutable. Any inferences about gays and lesbians drawn from this sample, however, are based on very small samples. For example, in a sample of 3,432 American men and women, only 12 women identified themselves as lesbians and only 27 men identified themselves as gay. As we discuss below, far more men and women than indicated by these figures have had same-sex experiences.

A second data source is the General Social Survey. To our knowledge, Badgett's (1995) study of earnings and sexual orientation, based on pooled 1989–1991 GSS data, was the first work exploiting the GSS to systematically compare gays and lesbians with heterosexual counterparts. For most of her analysis, Badgett (1995) defines lesbians, gays, and bisexuals as individuals having more same-sex sexual partners than opposite-sex sexual partners since age 18. Using this definition and conditioning on a variety of characteristics, she finds that gay men earn 28% less than hetero-

sexual men, but that sexual orientation has no statistically significant effect on women's earnings.

Black et al. (1998) provide a similar analysis using GSS data from 1988–1991, 1993, 1994, and 1996. They find that the effect of sexual orientation on earnings depends to some degree on the definition of sexual orientation that is used. One general finding in their work confirms Badgett's earlier finding for men: Gay men appear to earn substantially less than other men with equal skills. Lesbian women, however, earn 20% to 35% more than other equally skilled women; this difference is statistically significant.

Pooling the GSS samples over eight years produces a reasonable sample of gays and lesbians, between about 150 and 450 individuals (depending on the definition used), along with several thousand other men and women. Though obviously more satisfactory than the sample taken from the NHLS, this is still quite small.

A third data source, the 1990 U.S. census, allows a sample of more than 13,700 gays and lesbians to be identified. In taking the decennial census, the Census Bureau designates as the head of household (the householder) "the member (or one of the members) in whose name the home is owned, being bought, or rented." The Census Bureau then collects information on all the members of the household and identifies each member by his or her relationship to the householder. Before 1990, couples living outside marriage in marriage-like relationships were not identified separately from individuals living together as roommates. Demographers, however, had noticed an increasing prevalence of the former type of household. Bumpass and Sweet (1989), for example, report that only 3% of women born between 1940 and 1944 had ever cohabited by age 25; among women born 20 years later, 37% reported cohabiting by age 25. Because of this trend, the Census Bureau changed the survey instrument for the 1990 census to allow unmarried partners to be identified separately from roommates.

Fortunately the census instrument allows household heads to report an unmarried partnership regardless of the partner's sex. In contrast, many previous surveys (e.g., the National Longitudinal Study of the Class of 1972) explicitly restricted cohabitation questions to heterosexual partnerships. In the public use samples of the 1990 census, we can identify a sample of more than 6,800 gay and lesbian households. Clearly, this is not a random sample of people who would identify themselves as gay or lesbian, nor is it a sample of those who have engaged in same-sex sex, because the sample contains only individuals who are involved in a cohabiting relationship. Exploring the nature of this sample is a major contribution of this paper.

To our knowledge, the 1990 census data were first used to study a group of gays and lesbians by Lisa Krieger (1993), a reporter for the *San Francisco Examiner*. Since then, these data have proved useful for several academic studies. Black et al. (1997) examined the effects of sexual orientation on men's wages. They found that men in gay couples earn substantially less than other men, with controls for earnings-related characteristics such as potential experience, educa-

1. Kinsey's subjects were all recruited purposefully, not drawn from a known sampling frame. As is still common in research on gays, Kinsey selected his subjects from many venues, including institutions such as prisons and reform schools.

tion, and demographic traits (e.g., race). Much of this wage differential, however, may be explained by the occupational choices of the coupled gay men.

Klawitter (1997) studied the effects of sexual orientation on earnings among women. She found that women in lesbian couples earn substantially more than other women, but that much of this difference is attributable to differences in earnings-related characteristics. Klawitter and Flatt (1998), investigating the effects of state and local antidiscrimination policies for sexual orientation, found little evidence that these policies are correlated with higher earnings for gay men.

Black et al. (1999), who studied the geographic distribution of gay men, argued that gay men are more willing to pay for amenities not related to children, and provided evidence that this trait influences gay men to locate in unusually attractive locations. Jepsen (1998) and Jepsen and Jepsen (1999) studied assortative mating and labor market specialization of gay and lesbian couples.

Although these papers constitute a useful advance in understanding gays and lesbians in the United States, the reliability of their principal data source has not been investigated systematically.

IDENTIFYING GAYS AND LESBIANS IN SOCIAL SCIENCE DATA SETS

The General Social Survey (GSS) and the National Health and Social Life Survey (NHLS)

The GSS is designed to measure social indicators of opinions and attitudes over time in the United States. It uses a multistage area probability sampling design, in which a randomly selected adult from each household in selected geographic areas is asked to participate. Approximately 1,500 adults were sampled annually from 1972 through 1994 (except in 1981 and 1992). Since 1996, the sample size has been doubled but the survey has been conducted only every two years.

In 1992, when the GSS was not fielded, the resources were dedicated to fielding the NHLS. The NHLS was drawn using the same sampling frame as the GSS, and many of the same questions are asked, so that the two surveys can be combined for analysis. The NHLS is restricted to individuals ages 18–59; the GSS samples adults at any age over 18.

In comparing gays and lesbians with other men and women, one cannot avoid the complicated question of what it means to be gay or lesbian. The GSS and the NHLS contain a common set of questions on sexuality that allows several ways of defining sexual orientation. Beginning in 1988, the GSS has asked several questions about the sex of individuals with whom the respondent has had sex: Respondents were asked, "Have your sex partners in the last 12 months been exclusively male, both male and female, exclusively female?" Beginning in 1991, a parallel question has been asked about a respondent's sex partners in the five years prior to the survey. In each year since 1989 the GSS has asked both male and female respondents, "Now thinking about the time since your 18th birthday (including the past 12 months),

how many male partners have you had sex with?" A parallel question is asked about the number of female partners.

In combination with the respondent's sex, these questions can be used to classify a respondent's sexual orientation by four different definitions. The first and second definitions, "having ever had a same-sex sex partner" and "having had at least as many same-sex as opposite-sex sex partners since age 18," rely on information about the sex of sex partners since age 18. The third and fourth definitions, "having had exclusively same-sex sex over the last year" and "having had exclusively same-sex sex over the last five years," rely on the sex of sex partners over the last year or last five years. The first definition is employed in Laumann et al. (1994), Badgett (1995), and Black et al. (1998). The second definition is used in Badgett (1995) and Black et al. (1998). The third and fourth definitions are used in Laumann et al. (1994) and Black et al. (1998).

In Table 1 we report the incidence rates and sample counts of gay and lesbian status based on these four different definitions, using a sample that pools data from the 1989–1991, 1993, 1994, and 1996 GSS and NHLS. Like Laumann et al., we find that the incidence rate of homosexuality varies greatly depending on how homosexuality is defined. For example, 4.7% of men in the combined samples have had at least one same-sex experience since age 18, but only 2.5% of men have engaged in exclusively same-sex sex over the year preceding the survey. Similarly, 3.5% of women have had at least one same-sex sexual experience, but only 1.4% have had exclusively same-sex sex over the year preceding the survey. Table 1 also shows that regardless of definition, the samples of gays and lesbians are small in the GSS and NHLS, even when we combine seven years of data.

Table 2 shows that the definitions of homosexuality are not correlated as highly as one might think, particularly for women. For example, we find that among women who had at least one female sex partner since age 18, only 28% have been involved, over the past year, in exclusively same-sex sexual relationships. Similarly, only 42% of men who have had a male sexual partner since age 18 have had exclusively same-sex sex over the year before the survey.

The NHLS has two unique features that make it far more valuable than simply an additional year of GSS data. First, it is the only large probability survey that asks respondents directly about their sexual orientation. The questionnaire asks, "Do you think of yourself as heterosexual, homosexual, bisexual, or something else?" Table 1 shows that the incidence rate of homosexuality is slightly lower by this definition than by the definition of having had exclusively same-sex sex over the past year. Because of the low incidence rate (and the modest sample size of the NHLS), only 12 women and 27 men report thinking of themselves as homosexual.

A second unique feature of the NHLS is that the survey records detailed data on sexual partners and on living arrangements between the respondent and all sexual partners over his or her lifetime. As we make clear below, understanding partnership is crucial to understanding the sample of gays and lesbians identified in the 1990 census

TABLE 1. SAMPLE SIZES (AND INCIDENCE) OF GAYS AND LESBIANS FOR VARIOUS DEFINITIONS IN THE COMBINED 1988–1991, 1993, 1994, AND 1996, GSS AND NHLSLS^a

Definition of Homosexuality	Women		Men	
	Lesbian	Bisexual	Gay	Bisexual
1. At least one same-sex partner since age 18 ^b		260 (3.6%)		260 (4.7%)
Total Observations		7,125		5,536
2. More same-sex than opposite-sex partners since age 18		123 (1.8%)		164 (3.1%)
Total Observations		6,826		5,239
3. Same-sex and opposite-sex sex partners over the last year	—	29 (0.5%)	—	33 (0.6%)
Exclusively same-sex sex partners over the last year ^c	88 (1.4%)	—	139 (2.5%)	—
Total Observations		6,414		5,519
4. Same-sex and opposite-sex sex partners over the last 5 years	—	66 (1.2%)	—	72 (1.6%)
Exclusively same-sex sex partners over the last 5 years ^{c,d}	78 (1.5%)	—	115 (2.6%)	—
Total Observations		5,361		4,430
5. Self-identified gay, lesbian, or bisexual ^e	12 (0.6%)	10 (0.5%)	27 (1.8%)	11 (0.7%)
Total Observations		1,921		1,511

^aExcluded from the analysis are all individuals who showed an inconsistency suggesting that their recorded sex or sexual history might be in error. The GSS asked the respondent's sex and then the sex of each member of the household. If the GSS recorded different sexes for the respondent in these two parts of the survey, the observation was not used. In addition, if an individual indicated having had sex with a person of a particular sex over the last year (or last five years), but also reported never having had sex with a person of that sex since age 18, the observation was excluded from analysis. Similarly, if a respondent reported having had a child born to him or her, but reported never having had opposite-sex sex, the observation was dropped from analysis.

^bQuestions on the number of male and female partners of the respondent since age 18 were asked beginning with the 1989 survey.

^cCategories include exclusively same-sex sex, exclusively opposite-sex sex, sex with both men and women, and no sex during the relevant period. More individuals may have had exclusively same-sex sex over the last five years than over the last year, as more individuals have had sex at all over the last five years than over the last year.

^dAsked in the GSS since 1991.

^eAsked only in the NHLSLS.

data. In the NHLSLS data we can define gay and lesbian respondents as “partnered” if, at the time of the NHLSLS survey, they were cohabiting with a partner with whom there was a sexual relationship.

Unfortunately the GSS includes only limited information on sex partners and on cohabitation. In the GSS we know only the respondent's current household structure; identifying a household member as an “unmarried partner” is not an option. Since 1988, however, the GSS has asked respondents, “Was one of your partners (in the last 12 months) your husband, wife or regular sex partner?” In the GSS we can define a respondent as “partnered” if he or she either has a spouse in the household (for gays and lesbians, a spouse of the same sex), or lived with an unrelated adult in

the household and also reported having had sex with a “husband, wife or regular sex partner.”²²

2. This definition of partnership may be inaccurate if, for example, a gay respondent is living with a male roommate who is not his sexual partner. Unfortunately the GSS consistently reports only that there is an unrelated adult in the household; it does not consistently report whether this unrelated adult is a partner or a roommate. When a gay or lesbian person lives with an unrelated adult of the same sex, however, we find that the probability of that individual's having a regular sexual partner is close to 1. For example, of the 29 lesbians who were partnered by this definition, 27 had a regular sexual partner. (By comparison, fewer than half of the lesbians who were not partnered by this definition had a regular sexual partner.) Because of this ambiguity, we consider a gay or lesbian as partnered only if he or she lives with a person of the same sex identified as a spouse or an unrelated adult, and reports having a regular sexual partner.

TABLE 2. FRACTION OF MEN AND OF WOMEN WITH SAME-SEX SEXUAL EXPERIENCE WHO ARE GAY OR LESBIAN BY VARIOUS DEFINITIONS, COMBINED 1988–1991, 1993, 1994, AND 1996, GSS AND NHSLs

	2. More Same-Sex Than Opposite-Sex Sex Partners Since Age 18	3. Exclusively Same-Sex Sex Partners Over the Last Year	4. Exclusively Same-Sex Sex Partners Over the Last 5 Years
Men			
1. At least one same-sex sex partner since age 18	0.61 (n = 249)	0.42 (n = 209)	0.43 (n = 173)
2. More same-sex than opposite-sex sex partners since age 18		0.68 (n = 111)	0.69 (n = 98)
3. Exclusively same-sex sex partners over the last year			0.83 (n = 80)
Women			
1. At least one same-sex sex partner since age 18	0.42 (n = 233)	0.28 (n = 196)	0.21 (n = 170)
2. More same-sex than opposite-sex sex partners since age 18		0.51 (n = 78)	0.49 (n = 53)
3. Exclusively same-sex sex partners over the last year			0.66 (n = 44)

Source: Authors' compilations from the GSS-NHSLs data.

Notes: The denominator for each fraction is the number of individuals ever with a same-sex sexual experience since age 18 who could be classified as gay or lesbian by the relevant definition. For definitions based on sex of sex partners over the last year and over the last five years, the risk set consists of all individuals who had exclusively same-sex sex, who had exclusively opposite-sex sex, or who had sex with both men and women over the relevant period. Excluded are individuals who did not have sex over the relevant period, as well as individuals who did not answer the question. For the definition based on having had at least as many same-sex as opposite-sex sex partners, the risk set includes only individuals who reported an exact number of men and of women with whom they had had sex since age 18. People who refused to answer either question or who answered in a range (e.g., "more than one") are excluded.

Table 3 presents the partnership rates among men and women who had exclusively same-sex sex over the year preceding the survey. From the combined GSS and NHSLs data, we estimate that 28.4% of gay men and 44.1% of lesbians are partnered at the time of the survey. Then, using the NHSLs, with its rich cohabitation history, we find that 67.9% of gay men and 93.8% of lesbians lived with a same-sex sex partner at some time. Finally, we present the same statistics for the set of men and women in the NHSLs who self-identify as gays and lesbians. Although the samples are small, we find that the partnership rates among gays and lesbians who self-identify are quite similar to the rates for gays and lesbians defined by their sexual experience over the year before the survey.

Our first four definitions of sexual orientation in the GSS rely on accurate recording of the respondent's sex, as well as accurate recording of the sex of the respondent's partners. In general, accurate recording of sex is not an issue in social science data. Because only a small fraction of the U.S. population is gay or lesbian, however, inaccurate reporting of sex becomes a salient issue. To understand the issue at hand, consider an individual with one partner. Let the recorded sex of a respondent, *S*, and the recorded sex of a respondent's partner, *P*, take on one of two values, *M* or *F* (male or female). Let *S** and *P** be respectively the respondent's true sex and the true sex of a respondent's partner (that is, sex in the absence of recording error). We focus

on the following question: Given that there is inevitably some recording error in *S* and *P*, among men who were recorded "gay" (have same-sex partners), how many in fact are gay? A simple Bayes's rule calculation is helpful here:

TABLE 3. PARTNERSHIP RATES AMONG GAYS AND LESBIANS IN THE NHSLs AND GSS

	Sample Size	Percentage Currently Partnered	Percentage Ever Partnered
Same-Sex Sex Last Year (GSS and NHSLs)	161	34.2	NA
Gay	102	28.4	NA
Lesbian	59	44.1	NA
Same-Sex Sex Last Year (NHSLs Only)	44	34.1	77.3
Gay	28	28.6	67.9
Lesbian	16	43.8	93.8
Self-Identified Gay (NHSLs Only)	39	25.6	66.6
Gay	27	18.5	59.3
Lesbian	12	41.6	83.3

Source: Authors' compilations from the GSS-NHSLs data.

$$\Pr(S^* = M, P^* = M | S = M, P = M) = \frac{\Pr(S^* = M, P^* = M) \times \Pr(S = M, P = M | S^* = M, P^* = M)}{\sum_{s \in \{M, F\}} \sum_{p \in \{M, F\}} \Pr(S^* = s, P^* = p) \times \Pr(S = M, P = M | S^* = s, P^* = p)} \quad (1)$$

Suppose, for example, that a respondent's sex is recorded incorrectly with the same frequency as is the sex of a respondent's partner. If, in addition, incorrect recording of sex is independent of a respondent's sexual orientation, Eq. (1) reduces to

$$\Pr(S^* = M, P^* = M | S = M, P = M) = \frac{\Pr(S^* = M, P^* = M) \times \Pr(S = M | S^* = M) \times \Pr(P = M | P^* = M)}{\sum_{s \in \{M, F\}} \sum_{p \in \{M, F\}} \Pr(S^* = s, P^* = p) \times \Pr(S = M | S^* = s) \times \Pr(P = M | P^* = p)} \quad (2)$$

Evidence suggests that the fraction of gays in the population, $\Pr(S^* = M, P^* = M)$, is approximately 2.5%; the fraction of lesbians in the population, $\Pr(S^* = F, P^* = F)$, is approximately 1.5%; and the remaining 96% of the population is divided approximately evenly between heterosexual men, $\Pr(S^* = M, P^* = F)$, and heterosexual women, $\Pr(S^* = F, P^* = M)$. To complete our example, suppose that 0.5% of own and partner's sex is in error. Then Eq. (2) indicates that 16% of the sample classified as gay is not gay and that 24% of the sample classified as lesbian in fact is not lesbian.

No study has been conducted to validate the accuracy of the demographic variables in the GSS or NHLS.³ The internal consistency of the GSS data can be checked in several ways, however. For example, the survey asks each respondent his or her sex twice, once to collect the respondent's background characteristics and again to collect the relationship between persons living in the same household. For the 1988–1996 surveys, these two reports of sex agree in 96.13% of all cases in which the respondent was the head of household or a spouse; this percentage suggests an error rate of about 1.97%.⁴

Similarly, there are multiple reports of the sex of the respondent's sex partners. During the 1989–1996 surveys, for example, 4,105 respondents reported having sex with a man over the last year, and also reported the number of men with whom they had sex since age 18. Clearly, if the respondent reported having sex with a man in the last year, but then reported never having had sex with a man since age 18, one of the reports is in error. This occurred in only 0.8% (33) of the cases, yielding an error rate of 0.4%. If the rate at which sex is recorded incorrectly is 1.97% and the rate at which partner's

3. Tom Smith, director of the GSS, gave us this information. It seems almost impossible that own or partner's sex is recorded incorrectly in the NHLS. Many questions in the NHLS are sex-specific. Therefore, if sex was recorded incorrectly, many of the responses to questions would be nonsensical, an error unlikely to go unnoticed.

4. We limit our analysis to respondents who were recorded as the head of household or the spouse of the household head. If we assume that the two sex reports are independent and that error rates are symmetric, then the two reports disagree whenever one records the respondent's sex correctly and the other does not. That is, the error rate = $0.0387 = (1 - 0.96130) = 2 \times \Pr(P = M | P^* = F) \times (1 - \Pr(P = M | P^* = F))$, so $\Pr(P = M | P^* = F) = 0.0197$.

sex is recorded incorrectly is 0.4%, Eq. (2) suggests that 32% of respondents classified as "gay" in fact are not gay and that 44% of respondents classified as "lesbian" are not lesbian. Because of the nature of the recording error, most respondents coded mistakenly as gay are heterosexual women with their sex miscoded, and most respondents recorded mistakenly as lesbians are heterosexual men with their sex miscoded.

It is important to understand the implication of this misclassification. For example, women generally earn less than men in the labor market. Given this, it would be easy to infer that men recorded as gay in our sample earn less than other men, even if gay men's average earnings are the same as other men's. In drawing conclusions about differences between gay men and other men, or lesbian women and other women, it is paramount that this measurement issue be addressed.

There is no infallible method for addressing this problem with the GSS. One reasonable approach is to use only observations in which the two reports of the respondent's sex agree. A second approach is to attempt to combine the various measures of sexual orientation into one reliable measure. For example, Black et al. (1998) examine the robustness of their analysis of gays and lesbians as follows: They start with the observation that only a small fraction of heterosexual men and women ever have sex with a member of the same sex (as indicated in Table 1). Gay men, however, typically have had sex with a woman, and lesbian women typically have had sex with a man, at some time since age 18. Black et al. (1998) thus suggest that one way of limiting the intrusion of sex misclassification is to exclude all men recorded as currently "gay" who have not had sex with a woman at some time since age 18. (Similarly, they exclude all women recorded as lesbians who have not had sex with a man at some time since age 18.) The cost of such a procedure is that a gay man who has never experimented with opposite-sex sex will be excluded from the gay sample. The benefit is that virtually all women who have been coded mistakenly as men will be eliminated from the sample of gay men. (The only women in the gay sample would be women whose sex is miscoded and who have had same-sex sex. In the sample sizes used here, this would be a very small number of individuals, probably zero.)

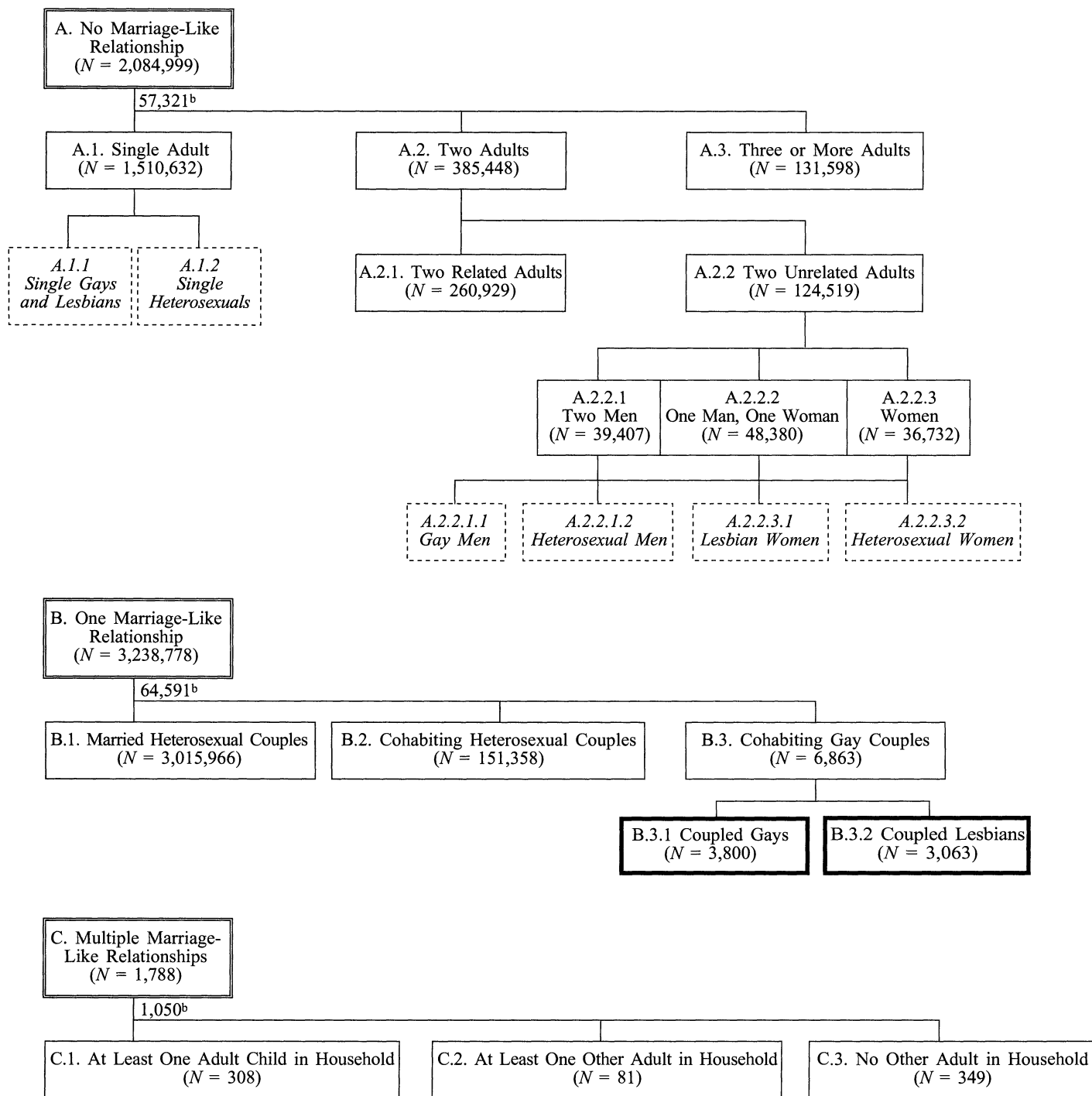
The 1990 Census Public Use Microdata Samples

Next we explore how the combination of the 5% and 1% Public Use Microdata Samples (PUMS) of the 1990 census can be used to construct a sample of gays and lesbians. In the 1990 census, as we have discussed, gay and lesbian respondents can identify themselves as unmarried partners (but not as married). These are the households we wish to study.

We begin with a sample of 6,632,090 households. Of these, we excluded 293,471 group quarters and 807,558 vacant housing units, as well as 205,494 households in which some household member's relationship to the householder was imputed.⁵ Figure 1 displays our scheme for classifying the remaining 5,325,565 households in the 1990 census.

5. We exclude households in which any member's relationship to the householder is allocated because this individual's exact relationship to the householder is unclear; as a result, the exact nature of the household struc-

FIGURE 1. CLASSIFICATION OF HOUSEHOLDS INTO MARRIAGE-LIKE RELATIONSHIPS FOR A SAMPLE WITH NONIMPUTED RELATIONSHIP TO HOUSEHOLDER FOR ALL HOUSEHOLD MEMBERS: PUBLIC USE MICRODATA SAMPLES, 1990 CENSUS^a



^aThe total number of households in the PUMS is 6,632,090. In 205,494 households, at least one household member had his or her relationship to the householder imputed by the Census Bureau. The sample contains 293,471 group quarters and 807,558 vacant households. Two households contain coding errors.

^bHouseholds with age allocated, sex allocated, or householder or partner under age 18.

The double-outlined boxes represent our three relationship categories: households in which the head has no apparent marriage-like relationship (A), households in which the head has a marriage-like relationship (B), and households in which the head appears to have multiple marriage-like relationships (C). A "marriage-like relationship" refers to a married couple or a partnered couple (either opposite- or same-sex). Each marriage-like relationship is divided into subgroups. Subgroups in boxes with solid outlines are, in principle, observable from the data, whereas subgroups in boxes with broken outlines are unobservable.

We find just over 2 million households with no marriage-like relationship. (We immediately exclude 57,321 of these households either because the householder was under 18 or because the householder's age was allocated.) The majority of these households with no marriage-like relationship are single adults (1.5 million). Obviously some of these individuals are gay or lesbian, but the census does not ask questions about respondents' sexual orientation, so they cannot be identified.

Of the remaining households, 131,598 are composed of three or more adults (A.3). Another 260,929 households consist of two related adults, which we also do not characterize further (A.2.1). Of the 124,519 households with two unrelated adults (A.2.2), 48,380 are of mixed sex (A.2.2.2), divided between opposite-sex roommates and cohabiting couples who chose not to identify themselves as unmarried partners, while 39,407 are two-male (A.2.2.1) and 36,732 are two-female (A.2.2.3) households. These households include heterosexual and homosexual individuals, but there is no means of identifying the respondents' sexual orientation.

About 3.2 million households contain one marriage-like relationship. (We immediately exclude 64,591 households with allocated age or sex and households in which either the partner or the householder was under 18.) Over 3 million are heterosexual married couples (B.1), 151,358 are opposite-sex couples who are unmarried partners (B.2), and, finally, 6,863 are same-sex unmarried partners: 3,800 male couples (B.3.1) and 3,063 female couples (B.3.2) (in boxes with bold outlines). These are same-sex couples with each partner over age 18. By excluding all households with any age allocation, any sex allocation, and any relationship-to-householder allocation, we can be assured that respondents indeed indicated that they were in a same-sex unmarried partnership.

Two important questions about our sample arise immediately. First, are same-sex partners identified in the sample

ture is also unclear. In doing this, we recognize that we also exclude some households with gay and lesbian partnerships in which an individual was identified as the "spouse" of a same-sex householder. Through a rather complex census recoding procedure, either the "spouse" in these same-sex couples was allocated into a different relationship status (such as roommate, unrelated adult, or possibly unmarried partner) or his/her sex was changed. In either case, the data in the PUMS do not allow the researcher to identify the specific reasons why an allocation of either relationship status or sex occurred. We are most grateful to Jason Fields, research analyst at the U.S. Census Bureau, for his assistance in researching the allocation process of same-sex spouses in the 1990 census.

indeed gay or lesbian, or instead are they nonpartnered individuals who were measured in error? Although the samples of gay and lesbian households are quite large, they account for only about 0.1% of all households in the census. Thus even small levels of misclassification of unmarried partnership status might lead to erroneous classification of a large fraction of same-sex couples in B.3.

Second, even if these 6,863 households are gay and lesbian, how do they relate to the population of gays and lesbians more generally? Obviously this sample necessarily excludes gays and lesbians not living in partnered relationships; furthermore, it is likely that the census greatly undercounts partnered gays and lesbians.

Our first concern is establishing a priori case that the same-sex partners we observe are not predominantly the product of measurement error. Below we discuss how partnered gays and lesbians differ from nonpartnered gays and lesbians, and how the census sample of partnered gays and lesbians differs from a random sample of partnered gays and lesbians drawn from the GSS.

One case that clearly shows some sort of measurement error is among households classified as containing multiple marriage-like relationships (C).⁶ The extent and type of error recorded in such cases provides an important clue about the extent and type of measurement error we might expect to face elsewhere in the data (particularly in B.3). The most prevalent reason for apparent multiple partnerships is that a child is classified as a partner or that a household member's age or sex was allocated. Of the 1,788 apparent multiple marriage-like relationships, 1,050 households show one of the partners under age 18 or with an allocated age or sex. This type of error will not affect our measurement of gay and lesbian couples because in all of our same-sex partnerships in B.3, both partners are 18 or older and all household members have nonallocated age and sex. There remain 738 multiple marriage-like relationships.

The first hypothesis we can investigate is that these errors are simply mistakes—cases in which a very small number of individuals randomly "checked the wrong box." If we examine households with exactly three adults, we find 86 households composed of one married couple and an additional partner. Altogether there are 450,717 three-adult households in which two of the adults are married. If we accept that spouses are identified correctly, then, since only 86 of 450,717 householders marked another adult as an unmarried partner, the error rate in recording "unmarried partner" among these households is trivial: less than 0.02%.

If we perform a similar exercise among households containing no married couples, the error rate is somewhat higher, but still very low. We find that among the 109,690 three-adult households without a marriage, 229 identify two of the adults as unmarried partners of the householder. In these cases, we believe the householder occasionally "checked the wrong box." This suggests a 0.2% error rate for this outcome.

6. We assume that polygamy or polyandry is negligible in our sample.

These calculations provide strong evidence that the error present among households with multiple marriage-like relationships does not take the form of random error ("checking the wrong box"), which affects all individuals with equal probability. Instead, a small but nontrivial number of individuals in households with no married couple apparently "misinterpret" the meaning of *unmarried partner*. For example, a respondent may consider his two roommates "partners" if household resources are shared. If this is the form of the error, it is easy to calculate its incidence. Altogether 153,048 households in our sample contain three or more adults but no married couples (this is the sum of the 131,598 households with no partnerships, 20,858 households with one partnership, 540 households with two unmarried partners recorded, and 52 with more than two unmarried partners). There are 592 errors among these households, for an error rate of 0.39%.

We consider even this low rate an upper bound on the rate of misclassification for our sample of gay and lesbian couples. Among the 592 errors by households with three or more adults but no married couple, 540 indicate that the householder has exactly two unmarried partners. We suspect that most of these cases are not the consequence of misinterpreting the meaning of *unmarried partner* per se; rather, we believe they entail the mistaken recording of relationships by a householder who lives with other adults, two of whom in fact are unmarried partners. The mistake made here by the householder is failure to recognize that he must mark his own relationship to each household member.

Further evidence of this type of error comes from the 308 households in which an adult child lives at home (C.1 in Figure 1). Inspection suggests that in many of these cases, the heterosexual unmarried partner of an adult child is recorded mistakenly as the householder's unmarried partner. (In the most common pattern, a husband and wife live with a child and an adult partner close in age to the child.) These sorts of error, however, are not likely to occur in our sample of same-sex couples, because only 15% of gay and lesbian couples live with any other adult in the household, and only 4% live with an adult child.

In sum, we read the evidence as suggesting that virtually none of the misclassification of "unmarried partner" status is due to random error or to confusion about the term; instead it is the result of mistakes concerning an individual's relationship to the householder. If in fact "unmarried partnership" was generally interpreted correctly, then misclassification is negligible when (as in the great majority of gay and lesbian couples) there are only two adults in the household. Even the worst possible case, however, in which about 0.4% of nonmarried householders mistake the meaning of *unmarried partner*, is not particularly serious for our analysis. Our sample contains just over 76,000 two-adult households in which both adults are of the same sex. If 0.4% of these householders mistakenly marked "unmarried partner," only about 305 couples would be identified mistakenly as gay or lesbian couples. In contrast, we actually identify nearly 7,000 cohabiting same-sex couples in our sample (B.3 in Figure 1).

As for sex miscoding, we can look to the 1970 census for validation. The Census Bureau studied the accuracy of the 1970 census data by matching the 1970 short- and long-form data to the 1970 Current Population Survey. Because many questions are the same in the CPS as on the census form, this matched file constituted a sample of over 20,000 cases. The Bureau reports that the error rate for sex among adults is less than 0.2% (U.S. Census Bureau 1975).⁷

Of course, a problem remains: The gays and lesbians in B.3 are only a fraction of all gays and lesbians in the population. One can calculate roughly how accurately the sample of same-sex partners identified in the census represents the gay and lesbian population in general. Suppose we adopt the reasonably narrow definition of *gay* and *lesbian* to be individuals who have engaged exclusively in same-sex sex over the last year. Then, according to Table 1, 2.5% of men are gay and 1.4% of women are lesbian. Given the estimated partnership rates for gays and lesbians of 28.4% and 44.1% respectively, we would estimate that in the United States, 0.71% of adult males are in gay-partnered households and 0.62% of women are in lesbian-partnered households. Our census sample contains 2,921,421 men age 18 to 60, of whom 7,287 are partnered gays, and 3,207,702 women age 18 to 60, of whom 5,762 are partnered lesbians. These latter statistics suggest that the number of households that self-report as same-sex couples in the 1990 U.S. census is considerably lower than the number counted in the GSS and NHLS samples. It would appear that roughly 35% of men living as partnered same-sex couples are recorded in the census; for women the corresponding fraction is 29%.

Given that only about one-third of cohabiting same-sex couples identify themselves as such in the census, extreme care is needed in drawing general inferences about the population of gay and lesbian couples. We certainly cannot rule out a priori the possibility that a householder's propensity to indicate that a same-sex partner is indeed an "unmarried partner" is correlated with individual characteristics such as age and education. We pursue one avenue for addressing this problem in the next section, comparing demographic characteristics of gay and lesbian couples in the census sample with corresponding information in the GSS and NHLS (which are much closer to true random samples). Other researchers may find it useful to employ the data to form bounds on parameters of interest.

Beyond this, researchers must use informed judgment as to the suitability of the census data for addressing particular questions. In estimating wage regressions, for example, Klawitter (1997) implicitly assumes that, after conditioning on a variety of characteristics (such as education and potential experience), self-identification by cohabiting lesbian women is not correlated with their wages in the labor mar-

7. No similar matched file exists for the 1990 census. In a second census study, the "Content Reinterview Survey," several thousand households from the 1990 long form roster were interviewed again six months after the original census survey to validate the 1990 census data. Although data on sex and partnership status were collected, the accuracy of these two data items is not reported in the Bureau's report (see U.S. Census Bureau 1993).

ket. We find this to be a reasonable working assumption, especially because the alternative is simply to abandon the important goal of studying the role of sexual orientation in labor market outcomes.

CHARACTERISTICS OF GAYS AND LESBIANS IN THE UNITED STATES: RESULTS FROM TWO DATA SOURCES

In examining the characteristics of the gay and lesbian population, we focus on four issues that are relevant for current discussions in public policy: the geographic distribution of gays and lesbians, military service, family structure, and earnings and earnings-related characteristics. As discussed above, great care is needed in interpreting results because gays and lesbians are defined differently in each data source. In addition, one must be sensitive to the likelihood that at least some persons in the sample of "gays" and "lesbians" are heterosexuals who have been miscoded (especially in the GSS). In this section we look for differences between gay men and other men, and between lesbian women and other women, that are qualitatively consistent across data sources. The presence of consistent differences between gays and lesbians and other individuals also helps establish the validity of the data sources.

Geographic Concentration

Our first use of the census sample is to provide some information about the geographic distribution of gay and lesbian couples in the United States. This information is interesting in its own right, and it is helpful also for providing further evidence that our sample of same-sex couples does not consist predominantly of misrecorded opposite-sex couples. If opposite-sex couples constituted the bulk of the sample, one would expect the geographic distribution of same-sex couples to be similar to the distribution of the population in the United States as a whole. This is very clearly not the case, however. In Table 4 we list the 20 cities in the United States with the largest populations of gay couples, and present a similar list for lesbian couples, along with the percentage of the total census sample contained in each city. The 20 cities with large numbers of gay couples, which are home to less than 26% of the U.S. population, contain nearly 60% of our sample of gay men. Clearly gay men are concentrated in a selected number of urban areas. Lesbian women are somewhat less geographically concentrated.

Comparison of columns (2) and (3) for men and of columns (5) and (6) for women shows that some cities have atypically high concentrations of gays and lesbians. For ex-

TABLE 4. 20 CITIES (PMSAs) WITH THE LARGEST GAY/LESBIAN-COUPLE POPULATIONS, 1990 CENSUS

Cities Ordered by Number of Gay Couples (1)	Percentage of Gay Sample in the City (2)	Percentage of U.S. Population in the City (3)	Cities' Orders by Number of Lesbian Couples (4)	Percentage of Lesbian Sample in the City (5)	Percentage of U.S. Population in the City (6)
1 Los Angeles, CA	9.77	3.57	New York, NY	6.03	3.39
2 New York, NY	8.37	3.39	Los Angeles, CA	5.35	3.57
3 San Francisco, CA	7.90	0.65	San Francisco, CA	3.27	0.65
4 Washington, DC	4.42	1.54	Minneapolis, MN	2.90	0.92
5 Chicago, IL	3.65	2.44	Washington, DC	2.84	1.54
6 Atlanta, GA	2.60	0.98	Seattle, WA	2.51	0.79
7 San Diego, CA	2.56	1.01	Boston, MA	2.48	1.08
8 Oakland, CA	2.54	0.84	Chicago, IL	2.47	2.44
9 Boston, MA	2.30	1.08	Oakland, CA	2.46	0.84
10 Seattle, WA	1.85	0.79	Philadelphia, PA	2.14	1.95
11 Dallas, TX	1.76	1.03	Sacramento, CA	1.69	0.59
12 Houston, TX	1.66	1.30	Atlanta, GA	1.55	0.98
13 Philadelphia, PA	1.61	1.95	San Diego, CA	1.53	1.01
14 Anaheim, CA	1.46	0.97	Baltimore, MD	1.39	0.95
15 Minneapolis, MN	1.31	0.92	Tampa, FL	1.33	0.83
16 Fort Lauderdale, FL	1.20	0.51	Portland, OR	1.31	0.47
17 Tampa, FL	1.17	0.83	Houston, TX	1.16	1.30
18 Phoenix, AZ	1.09	0.85	Phoenix, AZ	1.08	0.85
19 Denver, CO	1.07	0.63	Denver, CO	1.06	0.63
20 Sacramento, CA	1.04	0.59	San Jose, CA	0.93	0.60
Total	59.33	25.86	Total	45.48	25.38

Source: Authors' compilations from the 1990 U.S. census, 5% PUMS.

Note: Results in this table are drawn only from the 5% PUMS because the MSA definitions are not completely consistent across the 5% and 1% samples.

TABLE 5. PERCENTAGE VETERANS FOR PARTNERED GAYS/LESBIANS AND OTHER MEN/WOMEN, U.S. CENSUS

	Partnered Gay Men		Other Men		Partnered Lesbian Women		Other Women	
Military Service for Individuals Ages 18–67								
Any Military Service	17.3		36.8		6.6		1.4	
Veteran	15.5		31.1		5.1		1.0	
Reserves	1.6		4.2		1.2		0.2	
Military Service and Years Served for Cohorts Reaching Age 18 During Specific Eras	Percentage With Any Military Service	Average Number of Years Served	Percentage With Any Military Service	Average Number of Years Served	Percentage With Any Military Service	Average Number of Years Served	Percentage With Any Military Service	Average Number of Years Served
World War II Era (1941–1947)	72.7	3.0	75.0	3.5	10.2	1.6	1.6	0.8
Korean War Era (1950–1954)	60.6	3.6	64.3	3.9	11.9	3.5	1.0	0.6
Between Korean and Vietnam War Eras (1955–1964)	38.6	3.0	49.9	3.6	6.1	3.0	0.9	0.6
Vietnam War Era (1965–1974)	17.3	3.1	30.6	3.5	6.7	3.0	1.4	1.2
Post-Vietnam War Era (1975–1980)	6.9	2.6	14.6	3.7	7.2	2.8	1.9	1.4
Current Era (1981–1990)	7.0	1.7	14.0	2.6	5.1	1.4	1.7	1.0

Source: Authors' compilations from the 1990 U.S. census, 5% and 1% PUMS.

ample, a randomly selected gay man in our sample is about 12 times more likely to live in San Francisco than are other individuals in the U.S. population. Other cities with especially high concentrations of gays include Los Angeles, Washington, DC, and Atlanta. High concentrations of lesbian women are found in San Francisco, Seattle, and Minneapolis. When we look at concentrations of gay and lesbian couples in smaller cities (e.g., 200,000 to 700,000), we find a disproportionate number of “college towns” such as Ann Arbor and Madison. (For both gays and lesbians, seven of the 10 smaller cities with high concentrations contain a major university.) Inference about the gay and lesbian population in smaller cities, however, is severely limited by small sample sizes.

Unfortunately there exist no reliable data, other than the census, suitable for calculating even the most rudimentary statistics on the location of the gay and lesbian population. (We cannot report the geographic distribution of gays and lesbians from the GSS because this information is confidential.) Thus we cannot compare our results with findings from other samples.

We can gain some confirmatory evidence, however, by investigating the spatial distribution of AIDS deaths in 1990. In that year, an overwhelmingly large fraction of men who died of AIDS were gay.⁸ Using the 1990-detail mortal-

ity file (a complete enumeration of deaths in the United States), we calculated the total number of deaths and the number of deaths by AIDS for white men ages 25–44 in 300 SMSAs. Of 2,151,890 deaths in 1990, 74,600 occurred to white men ages 25–44; of these deaths, 12,844 were diagnosed as due to AIDS. For each of the identified SMSAs in the first panel of Table 4, we construct an AIDS death concentration ratio: the fraction of AIDS deaths in a city divided by the corresponding fraction of all deaths in the city. We find that this index is highly correlated with a “gay concentration index,” formed by dividing the figures in column (2) by the corresponding figures in column (3). In particular, among the 20 cities containing the largest numbers of gay men, the correlation between the gay concentration index and the AIDS death index is 0.89 (significantly different from zero at the 0.01 level). Fifty-nine percent of same-sex male partners live in the 20 cities listed; 54% of all deaths from AIDS to white men ages 25–44 occurred in these 20 cities.

Veteran Status

In regard to the interesting issue of military service among gays and lesbians, Table 5 provides for a comparison of mili-

AIDS in 1990 were infected as a result of same-sex male sexual contact alone or a combination of same-sex male sexual contact and injection drug use; 87.1% of those who subsequently died belonged to the same two risk categories; and 90% diagnosed with AIDS in the 1980s came from these two risk categories.

8. The following statistics are taken from the web site for the U.S. Centers for Disease Control and Prevention (<http://wonder.cdc.gov/aids00.shtml>), for white men aged 25–44: 86.7% who were diagnosed with

tary service for same-sex partners and other men and women. Although it appears that partnered gay men are much less likely than other men to be veterans, a substantial proportion (17.3%) of the sample are veterans, in the reserves, or on active duty in the military (compared with 36.8% for other men). For comparison, we conducted the same analysis with the GSS-NHSLs data and obtained similar results: 16.9% of gay men and 32.3% of heterosexual men are veterans. Among women, statistics based on census data show that 6.6% of the lesbian sample have served in the military, compared with 1.4% of other women. Again, estimates from the GSS-NHSLs data yield similar results: Military service rates are 8.1% for lesbian women and 1.4% for other women.⁹

The lower panel of Table 6 shows an interesting pattern. Gay men who reached draft age (age 18) during the World War II and Korean War eras served in the military at nearly the same rate as other men. In addition, during this period, gay men who were in the military served on average only slightly fewer years than other men. The rate of military service has been declining for men in general over the past several decades, and has decreased more rapidly for gay men than for other men. During the current era it appears that the fraction of gay men in the military is substantially lower than the fraction of other men, and that gay men are serving fewer years than other men.

Among women in same-sex partnerships who reached enlistment age (18) during the World War II and Korean War eras, over 10% served at some time in the military. Among other women in this cohort, 1.0% to 1.6% served at some time. Women in same-sex partnerships who served in the military accumulated far more years of duty than other women. In fact, until the post-Vietnam cohort, the number of years of service for lesbians who served in the military was similar to the number of years served by men. More recently, there has been a narrowing in the large differences between lesbian women and other women in military service rates, perhaps as a result of changes in military policy that previously barred married women and pregnant women from service.

Family Structure

Although adoption and parental rights policy for gay and lesbian couples is an intensely debated topic, we have virtually no empirical evidence regarding the current presence of children among gay and lesbian couples. The census sample provides the first reliable statistics on this matter. As shown in Table 6, these data indicate that a substantial number of same-sex couples, especially lesbian couples, currently have children present in the home: about 21.7% of partnered lesbians and 5.2% of partnered gays. Most of these children are relatively young: 71% of the children in lesbian households and 76% of those in gay households are under age 18. The

combined GSS-NHSLs data tell a more dramatic story. These data, which of course include single gays and lesbians, and gays and lesbians who are married (and may still live with their spouses), indicate that over 14% of gays and over 28% of lesbians have children in the household.

Many of the children in gay and lesbian households recorded in the census were probably born in previous marriages. In the census sample, nearly 20% of men in gay partnerships and 30% of women in lesbian partnerships were married previously or (in a small number of cases) are currently married. (See the far right-hand column of Table 7.) Among gays and lesbians more generally, as measured in the GSS-NHSLs, an even higher proportion are previously or currently married: possibly as many as 30% of the gay men and 46% of the lesbians (see Table 7).

Yet these measured rates decline somewhat when we make sample adjustments designed to reduce classification error. In particular, it seems very likely that a married gay or lesbian individual will have had opposite-sex experience. Thus readers may wish to focus on the third column in Table 7, which shows that about 25% of gay men and 40% of lesbian women are married or previously were married.

Education, Earnings, and Wealth

Education. The gays and lesbians in the census sample appear to be highly educated, span the distribution of ages, and are similar in racial makeup to the population as a whole. Table 8 indicates that same-sex partners generally have achieved higher levels of education than other individuals. (This is true of all the cohorts we examine, although in Table 8 we provide data only for ages 25–44.) The GSS-NHSLs data reveal a very similar pattern: For example, 13.0% of gay men have postcollege education and a further 23.7% have earned college degrees. The corresponding rates for married men are 10.3% and 17.0%. Among les-

TABLE 6. PRESENCE OF CHILDREN IN HOUSEHOLDS, U.S. CENSUS: CHILDREN AT HOME BY RELATIONSHIP STATUS (PERCENTAGES)

	Partnered Gay/ Lesbian	Partnered Heterosexual	Married	Not Partnered
Men				
No children	94.8	63.8	40.8	95.2
1 child	3.0	18.1	22.4	2.9
2 children	1.2	11.0	23.0	1.4
≥ 3 children	1.1	7.1	13.8	0.5
Women				
No children	78.3	63.8	40.8	77.9
1 child	12.6	18.1	22.4	10.1
2 children	5.0	11.0	23.0	7.6
≥ 3 children	4.1	7.1	13.8	4.5

Source: Authors' compilations from the 1990 U.S. census, 5% and 1% PUMS.

9. For the GSS-NHSLs calculations we conducted a "household roster check" to eliminate households for which sex identification is uncertain. We were left with quite small samples for gays ($n = 77$) and especially for lesbians ($n = 37$).

TABLE 7. CURRENT AND PAST MARITAL STATUS FOR GAYS AND LESBIANS, GSS (1988–1996) AND NHSLs (1992) DATA AND CENSUS DATA (PERCENTAGES)

Marital Status	GSS/NHSLs			Census
	Gay/Lesbian	Gay/Lesbian With Household Screening	Gay/Lesbian With Opposite-Sex Experience	Partnered Gay/Lesbian
Men				
Currently married	20.1	18.3	14.6	1.3
Widowed, separated, or divorced, and not currently married	11.6	11.5	11.1	17.2
Never married	68.4	70.2	74.6	81.4
(Sample size)	(139)	(131)	(63)	(7,567)
Women				
Currently married	23.9	13.9	6.5	1.2
Widowed, separated, or divorced, and not currently married	21.6	22.2	32.6	28.7
Never married	54.6	63.9	60.9	70.1
(Sample size)	(88)	(72)	(46)	(6,081)

Source: Authors' compilations from the GSS-NHSLs data and the 1990 U.S. census, 5% and 1% PUMS.

bian women, 13.9% have postcollege education and 25.0% have college education; comparable rates for married women are 6.1% and 16.0%.

The high educational levels of gays and lesbians in the census data may reflect poorly educated gays' and lesbians' relative unwillingness to indicate an unmarried partnership status on the census form; a similar selection bias may occur in the GSS-NHSLs data. Addressing this issue properly is a not a trivial matter. Here we pursue only one of the

potential avenues for examining this type of selection in the GSS.

It is well known that an individual's education is correlated with his parents' education. Suppose we take as a working assumption that the relationship between an individual's education and his father's education is the same for gay men as for other men. Now consider the null hypothesis that educational attainment is the same for gay men as for other men. (Under this null hypothesis, gay individuals' educational lev-

TABLE 8. EDUCATIONAL ATTAINMENT BY AGE AND RELATIONSHIP STATUS, U.S. CENSUS (PERCENTAGES)

Level of Education	Men				Women			
	Not Partnered	Heterosexual Partnered	Gay Partnered	Married	Not Partnered	Heterosexual Partnered	Lesbian Partnered	Married
Age 25–34								
Some high school	8.72	21.56	5.57	13.59	14.87	17.95	7.03	11.83
High school diploma	24.80	34.36	14.99	31.57	25.74	31.46	12.58	32.10
Some college	30.84	27.69	36.84	30.39	33.31	32.46	33.62	33.36
College degree	25.39	12.64	29.47	17.66	19.25	14.26	31.17	17.86
Postcollege	10.25	3.75	13.13	6.79	6.83	3.87	15.60	4.85
(Sample size)	(119,820)	(60,048)	(2,963)	(608,533)	(167,914)	(59,112)	(2,302)	(711,537)
Age 35–44								
Some high school	10.10	19.44	4.44	11.76	11.21	18.69	5.02	11.65
High school diploma	22.64	30.14	9.26	24.88	24.44	32.45	10.31	31.70
Some college	32.08	31.12	30.22	30.72	34.27	32.02	26.86	31.00
College degree	21.53	12.50	31.82	19.28	17.61	10.97	28.50	16.62
Postcollege	13.65	6.80	24.26	13.36	12.47	5.87	29.31	9.03
(Sample size)	(108,534)	(35,364)	(2,358)	(731,765)	(143,346)	(31,554)	(1,909)	(737,449)

Source: Authors' compilations from the 1990 U.S. census, 5% and 1% PUMS.

els are relatively high in the GSS because highly educated gay men are more willing than poorly educated gay men to report same-sex sexual relationships.) Given our working assumption, we would expect that the gay individuals in the GSS sample would have disproportionately well-educated fathers. Under the null hypothesis, the true distribution of father's education is the same for the gay population as for other men, but the observed distribution of father's education will be skewed toward higher levels of education because of selection bias (because fathers of poorly educated gay men are less likely to appear in the sample).

Empirically, however, we find no evidence of such a pattern. Instead we find that the distribution of education among gays' fathers is almost identical to that among other men's fathers in the GSS-NHSLs data. In turn, this finding provides tentative evidence that the gay men in fact accumulate more education than other men. (Similar analysis for women is hampered by small sample sizes.)

Earnings. Given the high levels of education among the gays and lesbians in both the census and the GSS-NHSLs data, it would not be surprising to find that gays and lesbians do relatively well in earnings. More important, there is evidence that sexual orientation affects earnings, even with conditioning on age and education. In Table 9 we present the

mean annual earnings by age category, educational level, and relationship status for the census data.

Two obvious empirical regularities are present. First, partnered gays earn substantially less than married men. Second, lesbian women earn substantially more than married women. Furthermore, lesbians in the census data generally earn more than single women and heterosexually partnered women. Patterns in the GSS-NHSLs data are remarkably similar. A parametric approach employed by Black et al. (1998) provides strong evidence that this pattern is statistically significant and robust to various definitions of sexual orientation. Black et al. suggest two possible interpretations for the observed effects of sexual orientation on earnings: One interpretation is based on Becker's (1981) model of specialization in the family, and the other is consistent with a subtle model of sex/sexual orientation discrimination.

Wealth: Homeownership. The census contains only limited information about household wealth. One important exception is homeownership. Two obvious patterns emerge in an analysis of homeownership and value by age category and relationship status. First, regardless of age category, the rate of homeownership is lower for partnered gays' and lesbians' households than for married-couple households. Second, conditional on owning a house, lesbian couples appear

TABLE 9. MEAN EARNINGS BY AGE, EDUCATION, AND RELATIONSHIP STATUS, U.S. CENSUS

Level of Education	Men				Women			
	Not Partnered	Heterosexual Partnered	Gay Partnered	Married	Not Partnered	Heterosexual Partnered	Lesbian Partnered	Married
Age 25-34								
Some high school	16,784	14,660	12,579	18,450	9,178	9,029	12,243	9,018
High school diploma	21,313	19,178	18,777	24,010	13,987	13,512	17,473	11,988
Some college	23,886	22,254	21,039	27,298	18,085	16,977	18,932	15,101
College degree	30,729	29,162	28,618	35,851	26,012	23,409	24,265	21,348
Postcollege	36,090	36,072	32,465	42,292	29,955	28,011	26,028	26,580
(Sample size)	(100,234)	(48,693)	(2,563)	(520,392)	(122,971)	(44,829)	(2,001)	(475,580)
Age 35-44								
Some high school	18,449	17,051	19,646	21,883	11,207	10,994	13,901	10,877
High school diploma	24,344	23,389	20,253	28,884	17,179	16,236	18,961	13,567
Some college	28,588	27,533	26,636	34,094	21,727	20,734	25,023	17,071
College degree	37,468	38,629	36,054	46,424	29,174	28,734	28,387	21,448
Postcollege	47,261	49,251	42,339	59,031	35,115	34,395	34,427	29,633
(Sample size)	(84,815)	(26,709)	(1,915)	(600,135)	(110,335)	(23,591)	(1,580)	(488,047)
Age 45-54								
Some high school	20,171	19,198	18,717	24,705	12,915	12,181	19,106	11,545
High school diploma	27,129	27,219	26,047	32,060	18,119	17,001	21,698	14,351
Some college	33,057	33,650	33,750	39,218	22,872	21,962	24,999	17,920
College degree	43,748	46,448	47,541	55,623	29,298	27,910	30,653	21,927
Postcollege	52,339	59,440	48,467	59,440	35,410	35,211	43,106	31,090
(Sample size)	(45,234)	(11,415)	(776)	(412,883)	(60,693)	(9,006)	(487)	(307,676)

Source: Authors' compilations from the 1990 U.S. census, 5% and 1% PUMS.

to have somewhat more expensive homes than their heterosexual counterparts, and gay couples appear to have much more expensive homes than their heterosexual counterparts. Consider, for example, individuals age 35–44: Sixty percent of partnered gay men own homes, and over 67% of those who do so have homes valued at \$100,000 or more. Similarly, about 65% of partnered lesbian women own homes, and 55% of these women have homes worth \$100,000 or more. By way of comparison, homeownership rates are about 80% for both married men and women, but only about 45% of these homeowners have homes valued at \$100,000 or more. In view of the differences in education levels we documented above, this finding simply may reflect the fact that gay and lesbian couples are more highly educated on average than heterosexuals (and thus have relatively high earnings). When we repeat the analysis for a set of individuals who have obtained a college degree, the magnitude of the differences between gay and lesbian couples and their heterosexual counterparts is reduced, but the basic pattern remains.

CONCLUSION

The three data sets that we examined here can be used by scholars to produce demographic research on the gay and lesbian population. The GSS and NHLS provide a fairly small random sample of gay and lesbian individuals in the United States. We document a number of measurement-error problems that can produce misclassification of sexual orientation by researchers using the GSS-NHLS, and we suggest some solutions. The 1990 U.S. census provides a much larger sample of gays and lesbians, but only partnered gays and lesbians can be studied. Our analysis of these data suggests that nearly all of the same-sex couples identified in the census data are in fact same-sex partners (not misclassified heterosexual couples). We estimate, however, that only about one-third of gay and lesbian couples report themselves as such in the census; thus selection bias is a potential concern.

Together these data allow us to offer a more complete statistical portrait of the gay and lesbian population than was previously available. Empirical observations along dimensions such as veteran status, education, and earnings are remarkably similar in the GSS-NHLS and the census. This fact increases confidence in the quality of the data sources and in the reported results in general. The findings of this paper include the following:

Sixty percent of partnered gay men in the United States are concentrated in only 20 cities, especially in cities such as San Francisco, Washington, DC, Los Angeles, Atlanta, and New York. A very high correlation exists between the cross-city distribution of gay couples, as measured in the 1990 U.S. census, and 1990 death rates of AIDS among young men; this correlation increases the face validity of the findings. Partnered lesbian women are somewhat less concentrated in large cities than are gay couples.

Gay men historically have served in the military in relatively large numbers. Gays who were young during the World War II and Korean War eras were about as likely as other

men to serve in the military. More recently, gay men have been less likely than other men to serve. In contrast, lesbian women are much more likely than other women to have served in the military, though this difference apparently has declined over the past five decades.

Many gay men and lesbian women have children. Non-negligible fractions of gay and lesbian people are currently married.

Gay and lesbian individuals have higher educational levels than other men and women. Suggestive evidence from the GSS-NHLS indicates that this finding is not the result of selection bias.

Gay men generally earn less than other men, whereas lesbian women generally earn more than other women. Future research on this topic may provide important clues about the nature of specialization in households, discrimination, and labor market outcomes.

Perhaps the most useful contribution of this paper is that it demonstrates the viability of doing credible empirical work on the gay and lesbian population with existing data sources. As the GSS provides additional waves and as the Census Bureau collects the 2000 census, considerably more data will become available for this purpose. (In addition, researchers have not yet exploited the confidential one-in-six files of the 1990 census for the study of the gay and lesbian population.)

The increasing body of research on gays and lesbians holds promise for improving the understanding of a population that previously has undergone little systematic research based on large samples. In addition, we believe, careful theoretical and empirical work that pays close attention to sexual orientation can help us to understand questions about the general nature of labor market choices, accumulation of human capital, specialization within households, and many other issues of interest to social scientists.

REFERENCES

- Badgett, M.V.L. 1995. "The Wage Effects of Sexual Orientation Discrimination." *Industrial and Labor Relations Review* 48: 726–39.
- Becker, G.S. 1981. *A Treatise on the Family*. Cambridge: Harvard University Press.
- Black, D.A., G. Gates, S. Sanders, and L. Taylor. 1997. "The Effects of Sexual Orientation on the Wages of Men." Working paper, Heinz School of Public Policy and Management, Carnegie Mellon University.
- . 1999. "Why Do Gay Men Live in San Francisco?" Working paper, Heinz School of Public Policy and Management, Carnegie Mellon University.
- Black, D.A., H. Makar, S. Sanders, and L. Taylor. 1998. "The Effects of Sexual Orientation on Earnings." Working paper, Heinz School, Carnegie Mellon University.
- Bumpass, L.L. and J. Sweet. 1989. "National Estimates of Cohabitation." *Demography* 26:615–25.
- Jepsen, L. 1998. *Essays on the Economics of Marriage: An Empirical Study of Same-Sex and Opposite-Sex Couples*. PhD dissertation, Department of Economics, Vanderbilt University.
- Jepsen, L. and C. Jepsen. 1999. "An Empirical Analysis of Same-

- Sex and Opposite-Sex Couples: Do 'Likes' Still Like 'Likes' in the '90s?" Working paper, Northwestern University.
- Kinsey, A.C., W.B. Pomeroy, C.E. Martin, and P.H. Gebhard. 1948. *Sexual Behavior in the Human Male*. Philadelphia: Saunders.
- Klawitter, M.M. 1997. "The Effects of Sexual Orientation on the Determinants of Earnings for Women." Working paper, Graduate School of Public Affairs, University of Washington.
- Klawitter, M.M. and V. Flatt. 1998. "The Effects of State and Local Antidiscrimination Policies on Earnings for Gays and Lesbians." *Journal of Policy Analysis and Management* 17:658-86.
- Krieger, L.M. 1993. "S.F. Remains a Mecca for Gay Couples: Census Data Reveal Same-Sex Households Abound in the City." *San Francisco Examiner*, September 12, pp. A1, A10.
- Laumann, E.O., J.H. Gagnon, R.T. Michael, and S. Michaels. 1994. *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago: University of Chicago Press.
- Michael, R.T., J.H. Gagnon, E.O. Laumann, and G. Kolata. 1994. *Sex in America: A Definitive Study*. Boston: Little, Brown.
- "S.F. Seeks Equal Treatment." 1996. *San Jose Mercury News*, December 29, p. 3b.
- U.S. Census Bureau. 1975. "Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match." *1970 Census of Population and Housing, Evaluation and Research Program Report*. Washington, DC: U.S. Government Printing Office.
- . 1993. "Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview." *1990 Census of Population and Housing, Research and Evaluation Report, CPH-E-1*. Washington, DC: U.S. Government Printing Office.